

AD-A060 387

CALIFORNIA UNIV DAVIS DEPT OF MATHEMATICS
PSEUDO MAXIMUM LIKELIHOOD ESTIMATION: THEORY AND APPLICATIONS.(U)
AUG 78 G G HANNON, F J SAMANIEGO

F/G 12/1

UNCLASSIFIED

TR-7

AFOSR-TR-78-1416

AFOSR-77-3180

NL

1 OF 1
AD
AO 60 387



TECHNICAL REPORT SERIES
ON
SIGNAL PLUS NOISE MODELS

D D C
REGINA



LEVEL II

2

AD A060387

DDC FILE COPY

6

Pseudo Maximum Likelihood Estimation:
Theory and Applications,

10

Gail G./Hannon² and Francisco J./Samaniego

9

Technical Report, No. 7

17

14 TR-7

16

23 & 4 / A5

18 AFOSR

11

August, 1978

19 TR-78-1426

Department of Mathematics
University of California
Davis, California

12 54p.

ABSTRACT for	
DTIC	White Section <input checked="" type="checkbox"/>
DDC	Buff Section <input type="checkbox"/>
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	AVAIL. and/or SPECIAL
A	

DDC
RECEIVED
OCT 26 1978
D

15

¹ This research was sponsored in part by the Air Force Office of Scientific Research under Contract AFOSR-77-3180.

² Now at Standord University.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

78 10 16 125

401 891 LB

SUMMARY

Let X_1, \dots, X_n be i.i.d. random variables with probability distribution $F_{\theta, p}$ indexed by two real parameters. Let $\hat{p} = \hat{p}(X_1, \dots, X_n)$ be an estimate of p other than the maximum likelihood estimate, and let $\hat{\theta}$ be the solution of the likelihood equation $\partial/\partial\theta \ln L(\underline{x}, \theta, \hat{p}) = 0$ which maximizes the likelihood. We call $\hat{\theta}$ a pseudo maximum likelihood estimate of θ , and give conditions under which $\hat{\theta}$ is consistent and asymptotically normal. Pseudo maximum likelihood estimation easily extends to k parameter models, and is of interest in problems in which the likelihood surface is ill-behaved in higher dimensions but well-behaved in lower dimensions. We examine several signal plus noise or convolution models which exhibit such behavior and satisfy the regularity conditions of the asymptotic theory. For specific models, a numerical comparison of asymptotic variances suggests that a pseudo maximum likelihood estimate of the signal parameter is uniformly more efficient than estimators that have been advanced by previous authors. A number of other potential applications are noted.

I. INTRODUCTION

Probability models abound for which the analytical derivation of the maximum likelihood estimate of model parameters is virtually impossible. For many such models, one among the wide variety of numerical algorithms available for approximating the MLE will prove satisfactory. For other models, numerical methods are unreliable or converge too slowly to be of use. The recent paper by Dempster, Laird and Rubin (1977), and the associated discussion, are indicative of the fact that numerical procedures for approximating MLE's are still being vigorously investigated.

Among models whose elusive behavior with regard to the maximum likelihood approach has been studied extensively is the Cauchy model (see Barnett (1960) and Ferguson (1978)). There has also been difficulty with the likelihood approach to estimation in the presence of nuisance parameters. Neyman and Scott (1948) constructed an example of an inconsistent MLE in such a problem. Godambe (1977) has referred to this latter area as the major failure of the likelihood approach, and has developed the theory of estimating equations in part to fill this void. Godambe and Thompson (1974) treat optimal estimating equations in the presence of nuisance parameters. Kiefer and Wolfowitz (1956) proved the consistency of the maximum likelihood estimate in a particular version of the nuisance parameter problem, but it remains true that closed form expressions for the MLE in such problems are rare, and the multiplicity of solutions to the likelihood equations often renders numerical methods impractical.

The difficulties in obtaining the MLE in problems with nuisance parameters has led to the investigation of alternative estimation procedures which have the spirit of likelihood procedures, but are compromises due to the untractability of the preferred approach. Notable among such procedures are maximum integrated likelihood estimates, maximum marginal likelihood estimates and maximum conditional likelihood estimates (also called conditional MLE's). These three techniques are discussed in the paper by Kalbfleisch and Sprott (1970). The first of these has a Bayesian flavor in that one first eliminates the nuisance parameters by integrating the likelihood function with respect to some probability distribution on the appropriate portion of the parameter

space, and then obtains an estimate by maximizing the integrated likelihood. The maximum marginal likelihood estimate may be defined when the likelihood function factors into two components, one of which depends only on the "structural parameter" and not on nuisance parameters. When such a factorization is possible, the estimate is obtained by maximizing this latter component of the likelihood. The maximum conditional likelihood estimate is defined when there exists a statistic T with the property that the conditional distribution of the sample, given T , does not depend on the nuisance parameters. The estimate is obtained by maximizing the likelihood conditioned on T . These approaches have proven tractable in a number of problems which have resisted the direct maximum likelihood approach. Moreover, optimality results have been established for some of these procedures. For example, Andersen (1970) has shown that maximum conditional likelihood estimators are strongly consistent under regularity conditions, and has derived the parameters of their asymptotic (normal) distribution. Godambe (1976) gives conditions under which the conditional likelihood equation is the optimal estimating equation in a fixed sample size problem.

A comprehensive review of the literature on estimation in the presence of nuisance parameters is given in two papers by Basu ((1975) and (1977)). To the extent that approaches to estimation in nuisance parameter problems have focused on the elimination of nuisance parameters through conditioning or data reduction, the approaches have limited applicability. Many problems of practical importance do not give rise to convenient factorizations or to the existence of useful sufficient or ancillary statistics. The convolution models considered in the latter half of this paper are models

in which the approaches mentioned in the preceeding paragraph fail. It is precisely the characteristics of these models that has led us to the approach studied in this paper.

As a motivating example, consider a random variable X which presents itself as the sum of independent components Y and Z , to be referred to as signal and noise variables respectively. For concreteness, suppose Y is Poisson distributed with parameter θ and Z is binomially distributed with parameters (N,p) , with N assumed known. Signal plus noise variables such as X will be discussed in detail in Section III of this paper, and we postpone until then any discussion of the appropriateness of these models in describing random phenomena or of their stochastic and statistical properties. Let us focus on the problem of estimating θ , the signal parameter, assuming it to be the parameter of primary interest. A natural approach to this estimation problem is to estimate the pair (θ,p) and advance the estimate of θ so derived as appropriate. For the Poisson-binomial convolution, for example, one can estimate (θ,p) by the method of moments, as in done by Sclove and Van Ryzin (1969), and thus obtain an estimate of θ . Maximum likelihood estimation has not as yet been accomplished for this model, primarily due to the cumbersome nature of the likelihood function, which consists of the product of sums of products of component probabilities. Moreover, it is difficult to obtain the MLE numerically for this model since the likelihood surface has several local maxima. In the problem at hand, maximum likelihood estimation is untractable and the method of moments is inefficient. Moreover, neither method recognizes the special role played by the signal parameter θ , but, instead, treats both parameters equally. We study here an alternative

approach, that of pseudo maximum likelihood estimation, which focuses on the "structural" parameters of the model. For the Poisson-binomial convolution, for example, we estimate the noise parameter p by the method of moments, and, treating this estimate as the true value of p , estimate θ by maximizing the likelihood. Results established in this paper will imply that this pseudo MLE $\hat{\theta}$ is consistent and asymptotically normal. In a numerical comparison of the asymptotic variances of $\hat{\theta}$ and the method of moments estimator of θ , we find $\hat{\theta}$ to be uniformly superior.

In general, pseudo maximum likelihood estimation consists of replacing all nuisance parameters in a model by estimates and solving a reduced system of likelihood equations. The method is a reasonable one in problems in which lower dimensional maximum likelihood estimation is feasible while higher dimensional maximum likelihood estimation is untractable. As we shall see, the method is in this sense ideally suited for application to many convolution models. In Section II, we develop the asymptotic theory of pseudo maximum likelihood estimators, establishing under regularity conditions consistency and asymptotic normality. In Section III, some known results are summarized and some new results are obtained for several families of signal plus noise distributions. In particular, lower dimensional maximum likelihood estimation is shown to be feasible for a variety of such distributions. In Section IV, several signal plus noise distributions are shown to satisfy the regularity conditions under which the asymptotic theory of pseudo maximum likelihood estimates is developed. In Section V, we discuss the asymptotic relative efficiency of pseudo maximum likelihood estimates, and make some concluding general remarks in Section VI.

II. ASYMPTOTICS

Let X_1, \dots, X_n be a random sample from a member of the two parameter family $\mathcal{F} = \{F_{\theta, \pi}\}$ of distributions on the real line. We will assume throughout our development the existence, for every (θ, π) , of the density or probability mass function $f(x|\theta, \pi)$ with respect to some sigma-finite measure μ on \mathbb{R} . The method of pseudo maximum likelihood estimation may be viewed as follows. Given a sample of size n from $F_{\theta, \pi}$, an estimate $\hat{\pi}_n$ is developed for the parameter π by some technique or approach other than maximum likelihood estimation. The pseudo MLE is then obtained by maximizing the log likelihood $\mathcal{L}_n(\theta, \hat{\pi}_n)$, viewed as a function of the single parameter θ . The pseudo MLE $\hat{\theta}_n$ should have good large sample properties when $\hat{\pi}_n$ does. For example, if $\hat{\pi}_n$ is a consistent estimate of π and \mathcal{L} is a smooth function of (θ, π) near the true parameter value, then for large n the MLE and the pseudo MLE should be close with high probability. The consistency of the pseudo MLE is thus expected in smooth problems, and is established here under simple and natural regularity conditions. The efficiency of $\hat{\theta}_n$ will of course depend on the relative efficiency of $\hat{\pi}_n$. The asymptotic distribution of $\hat{\theta}_n$ is derived under regularity conditions for problems in which $\hat{\pi}_n$ is \sqrt{n} -consistent and asymptotically normal. The asymptotic theory for pseudo MLE's is developed here for a two-parameter problem rather than more generally because of the resultant ease of exposition and simplicity of notation. We trust that the validity of the extension to k -parameter problems will be apparent to the reader.

We will make use of the Mann-Wald symbols for convergence and boundedness in probability. Specifically, let $\{X_n\}$ be a sequence of

random variables and $\{a_n\}$ a sequence of positive constants. We write

$$X_n = O_p(a_n) \quad (2.1)$$

if $|x_n|/a_n \rightarrow 0$ in probability. The consistency of $\{X_n\}$ as estimates of θ may be written $(X_n - \theta) = O_p(1)$. We will write

$$X_n = O_p(a_n) \quad (2.2)$$

if $\forall \epsilon > 0$, there exists an integer N_ϵ , and $k_\epsilon \in (0, \infty)$, such that $\forall n > N_\epsilon$,

$$P\left(\frac{|X_n|}{a_n} < k_\epsilon\right) > 1 - \epsilon.$$

If $X_n = O_p(a_n)$, we say that $|X_n|/a_n$ is bounded in probability, and we use the phrase " X_n is \sqrt{n} -consistent for θ " to mean $\sqrt{n}(X_n - \theta) = O_p(1)$.

We will make use of the well-known result that if $X_n = O_p(a_n)$ and $Y_n = O_p(b_n)$, then $X_n Y_n = O_p(a_n b_n)$.

We make repeated use of a fundamental lemma which is elementary and perhaps well known. We have not seen it in the form below and present it with proof both for completeness and for emphasis.

Lemma 2.1. Let X_1, \dots, X_n be i.i.d. random variables from a distribution F_π on the real line, with $\pi \in \Pi \subseteq \mathcal{R}$. Let $\pi_0 \in \Pi$ be the true value of the parameter, and let $\hat{\pi}_n = \hat{\pi}_n(X_1, \dots, X_n)$ be such that $\hat{\pi}_n \xrightarrow{\text{Pr}} \pi_0$. Let $\psi(x, \pi)$ be a differentiable function of π for $\pi \in B$, an open neighborhood of π_0 , and for almost all x in the sample space \mathcal{X} , and suppose $E|\psi(X_1, \pi_0)| < \infty$. If

$$\left| \frac{\partial}{\partial \pi} \psi(x, \pi) \right| \leq M(x) \quad (2.3)$$

for all $\pi \in B$, where $EM(X_1) < \infty$, then

$$\frac{1}{n} \sum_{i=1}^n \psi(X_i, \hat{\pi}_n) \xrightarrow{\text{Pr}} E \psi(X_1, \pi_0). \quad (2.4)$$

Proof. Consider the Taylor series expansion

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \psi(X_i, \hat{\pi}_n) &= \frac{1}{n} \sum_{i=1}^n \psi(X_i, \pi_0) \\ &+ (\hat{\pi}_n - \pi_0) \cdot \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \pi} \psi(X_i, \tilde{\pi}_n) + o_p(1) \end{aligned} \quad (2.5)$$

where $\tilde{\pi}_n$ is between π_0 and $\hat{\pi}_n$. By the weak law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n \psi(X_i, \pi_0) \xrightarrow{\text{Pr}} E \psi(X_1, \pi_0).$$

The second term on the right of (2.5) is $o_p(1)$, which may be seen as follows. We have $\hat{\pi}_n - \pi_0 = o_p(1)$. Let $\epsilon > 0$, $\delta > 0$, and define

$$S_n = \{\hat{\pi}_n \in B\}$$

and

$$T_n = \left\{ \frac{1}{n} \sum_{i=1}^n M(X_i) < EM(X_1) + \delta \right\}.$$

Since $\hat{\pi}_n \xrightarrow{\text{Pr}} \pi_0 \in B$ and $\frac{1}{n} \sum_{i=1}^n M(X_i) \xrightarrow{\text{Pr}} EM(X_1)$, we may find

$N_\epsilon \in (0, \infty)$ such that $\forall n > N_\epsilon$, $P(S_n) > 1 - \epsilon/2$ and $P(T_n) > 1 - \epsilon/2$. We then have $\forall n > N_\epsilon$,

$$\begin{aligned}
& P\left(\left|\frac{1}{n} \sum \frac{\partial}{\partial \pi} \psi(X_i, \tilde{\pi}_n)\right| \leq E M(X_1) + \delta\right) \\
& \leq P(S_n \cap T_n) \\
& < 1 - \epsilon.
\end{aligned}$$

Thus, $\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \pi} \psi(X_i, \tilde{\pi}_n) = O_p(1)$ and $\frac{1}{n} \sum_{i=1}^n \psi(X_i, \hat{\pi}_n) - E \psi(X_1, \pi_0) = O_p(1)$, completing the proof.

We remark that the condition $E|\psi(X_1, \pi_0)| < \infty$ in Lemma 2.1 may obviously be weakened to $E\psi(X_1, \pi_0)^+ < \infty$. We now define the notation to be used in the sequel. Let X_1, \dots, X_n be iid, each with density or probability mass function $f(\cdot | \theta_0, \pi_0)$ defined on the sample space $\mathcal{X} \subseteq \mathcal{R}$, where $\theta_0 \in A$, A open with $A \subseteq \Theta \subseteq \mathcal{R}$ and $\pi_0 \in B$, B open with $B \subseteq \Pi \subseteq \mathcal{R}$. Let

$$\Phi(x | \theta, \pi) \equiv \ln f(x | \theta, \pi)$$

$$\mathcal{L}_n(\theta, \pi) \equiv \ln \sum_{i=1}^n f(x_i | \theta, \pi) = \sum_{i=1}^n \Phi(x_i | \theta, \pi)$$

$$\bar{\mathcal{L}}_n(\theta, \pi) \equiv \frac{1}{n} \mathcal{L}_n(\theta, \pi).$$

We will occasionally suppress the subscript n in the latter expressions. Partial derivatives are denoted with subscript notation; for example,

$$\Phi_\theta(x | \theta, \pi) \equiv \frac{\partial}{\partial \theta} \Phi(x | \theta, \pi)$$

and

$$\Phi_{\theta\pi}(x | \theta, \pi) \equiv \frac{\partial^2}{\partial \pi \partial \theta} \Phi(x | \theta, \pi)$$

(A4) For all $(\theta, \pi) \in A \times B$ and for all $x \in \mathcal{X}$,

$$\left| \frac{\partial}{\partial \pi} \log \frac{f(x|\theta, \pi)}{f(x|\theta_0, \pi)} \right| \leq M(x, \theta)$$

where $EM(X_1, \theta) < \infty \quad \forall \theta \in A$.

(A5) The following third partial derivatives are bounded by integrable functions:

$$(i) \quad |\phi_{\theta\theta\theta}(x|\theta, \pi)| \leq M(x) \quad \forall (\theta, \pi) \in A \times B, \forall x$$

$$(ii) \quad |\phi_{\theta\theta\pi}(x|\theta_0, \pi)| \leq M(x) \quad \forall \pi \in B, \forall x$$

$$(iii) \quad |\phi_{\theta\pi\pi}(x|\theta_0, \pi)| \leq M(x) \quad \forall \pi \in B, \forall x,$$

where $EM(X_1) < \infty$.

(A6) For any $(\theta, \pi) \neq (\theta_0, \pi_0)$,

$$P_{\theta_0, \pi_0} \{f(X_1|\theta, \pi) = f(X_1|\theta_0, \pi_0)\} < 1.$$

We first establish the consistency of the pseudo MLE.

Theorem 2.1. Let $X_1, \dots, X_n \stackrel{iid}{\sim} F_{\theta_0, \pi_0}$, and let $\hat{\pi}_n = \hat{\pi}_n(X_1, \dots, X_n)$

be a consistent estimate of π_0 . Under regularity conditions (A1), (A4) and (A6), the equation

$$\frac{\partial}{\partial \theta} \mathcal{L}_n(\theta, \hat{\pi}_n) = 0$$

has a solution $\hat{\theta}_n$ such that $\hat{\theta}_n \xrightarrow{\text{Pr}} \theta_0$.

Proof: Let $\theta \in A$ be fixed, and define

$$\begin{aligned}\psi(x|\pi) &= \Phi(x|\theta, \pi) - \Phi(x|\theta_0, \pi) \\ &= \ln \frac{f(x|\theta, \pi)}{f(x|\theta_0, \pi)}.\end{aligned}$$

By (A1), $\psi(x|\pi)$ is differentiable for $\pi \in B$, and by (A4),

$$|\psi_\pi(x|\pi)| \leq M(x) \quad \forall \pi \in B, \forall x.$$

Thus, by lemma 2.1,

$$\begin{aligned}\bar{J}_n(\theta, \hat{\pi}_n) - \bar{J}_n(\theta_0, \hat{\pi}_n) &= \frac{1}{n} \sum_{i=1}^n \psi(X_i, \hat{\pi}_n) \xrightarrow{\text{Pr}} E_{\theta_0, \pi_0} \psi(X_1 | \pi_0) \\ &= E_{\theta_0, \pi_0} \ln \frac{f(X_1 | \theta, \pi_0)}{f(X_1 | \theta_0, \pi_0)}\end{aligned} \quad (2.6)$$

It is easy to show that $E_{(\theta_0, \pi_0)} \psi(X_1 | \pi_0)^+ < \infty$ in general, and we thus claim that $\bar{J}_n(\theta, \hat{\pi}_n) - \bar{J}_n(\theta_0, \hat{\pi}_n)$ converges to a negative number (possibly $-\infty$) if $\theta \in A - \{\theta_0\}$. This claim follows from Jensen's inequality and condition (A6) since

$$E_{\theta_0, \pi_0} \ln \frac{f(X_1 | \theta, \pi_0)}{f(X_1 | \theta_0, \pi_0)} < \ln E_{\theta_0, \pi_0} \frac{f(X_1 | \theta, \pi_0)}{f(X_1 | \theta_0, \pi_0)} = 0.$$

Because of the convergence demonstrated above, we may find, for any $\epsilon, \delta > 0$ for which $(\theta_0 - \epsilon, \theta_0 + \epsilon) \subseteq A$, an integer $N_{\epsilon, \delta}$ such that $n > N_{\epsilon, \delta}$ implies that

$$P(\bar{\mathcal{L}}_n(\theta_0 - \epsilon, \hat{\pi}_n) < \bar{\mathcal{L}}_n(\theta_0, \hat{\pi}_n)) > 1 - \frac{\delta}{3}$$

$$P(\bar{\mathcal{L}}_n(\theta_0 + \epsilon, \hat{\pi}_n) < \bar{\mathcal{L}}_n(\theta_0, \hat{\pi}_n)) > 1 - \frac{\delta}{3}$$

and $P(\hat{\pi}_n \in B) > 1 - \frac{\delta}{3}.$

Thus, for $n > N_{\epsilon, \delta},$

$$P(\mathcal{L}_n(\theta, \hat{\pi}_n) \text{ has a local maximum } \hat{\theta}_n \in (\theta_0 - \epsilon, \theta_0 + \epsilon)) > 1 - \delta.$$

By (A1), $\hat{\theta}_n$ is a solution of the equation

$$\frac{\partial}{\partial \theta} \mathcal{L}_n(\theta, \hat{\pi}_n) = 0.$$

This completes the proof.

The usual remark about consistency applies here, that is, the above result does not imply that the pseudo MLE is consistent, but only that the pseudo maximum likelihood equation has a consistent root. In all applications considered in this paper, however, we are able to show that the pseudo maximum likelihood equation has a unique solution, so that in the models we consider, the pseudo MLE is indeed consistent. We turn to the asymptotic distribution of the pseudo MLE.

Theorem 2.2. Let $X_1, \dots, X_n \stackrel{iid}{\sim} F_{\theta_0, \pi_0}$, and let $\hat{\pi}_n = \hat{\pi}_n(X_1, \dots, X_n)$ be such that $(\hat{\pi}_n - \pi_0) = O_p(1/\sqrt{n})$, and suppose

$$\sqrt{n} \begin{bmatrix} \bar{J}_{\theta}(\theta_0, \pi_0) \\ \hat{\pi}_n - \pi_0 \end{bmatrix} \xrightarrow{d} N(0, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{22} \end{bmatrix}) \quad (2.7)$$

Then, under regularity conditions (A1)-(A6), the pseudo MLE $\hat{\theta}_n$ is asymptotically normal, that is

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \sigma^2) \quad (2.8)$$

where

$$\sigma^2 = \frac{1}{J_{11}} + \frac{J_{12}^2}{J_{11}^2} (\Sigma_{22} J_{12} - 2\Sigma_{12}) . \quad (2.9)$$

Proof. By condition (A1) and the consistency of $\hat{\theta}_n$ and $\hat{\pi}_n$, we may expand $\bar{J}_{\theta}(\hat{\theta}_n, \hat{\pi}_n)$ about θ_0 as follows:

$$\begin{aligned} 0 &= \sqrt{n} \bar{J}_{\theta}(\hat{\theta}_n, \hat{\pi}_n) = \sqrt{n} \bar{J}_{\theta}(\theta_0, \hat{\pi}_n) \\ &\quad + \sqrt{n}(\hat{\theta}_n - \theta_0) \bar{J}_{\theta\theta}(\theta_0, \hat{\pi}_n) \\ &\quad + \frac{1}{2} \sqrt{n}(\hat{\theta}_n - \theta_0)^2 \bar{J}_{\theta\theta\theta}(\tilde{\theta}_n, \hat{\pi}_n) + o_p(1), \end{aligned}$$

where $\tilde{\theta}_n$ lies between θ_0 and $\hat{\theta}_n$. Then

$$\begin{aligned}
-\sqrt{n} \bar{\mathcal{I}}_{\theta}(\theta_0, \hat{\pi}_n) &= \sqrt{n}(\hat{\theta}_n - \theta_0) \left[\bar{\mathcal{I}}_{\theta\theta}(\theta_0, \hat{\pi}_n) \right. \\
&\quad \left. + \frac{1}{2}(\hat{\theta}_n - \theta_0) \bar{\mathcal{I}}_{\theta\theta\theta}(\tilde{\theta}_n, \hat{\pi}_n) \right] + o_p(1),
\end{aligned}$$

which we rewrite as

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{-\sqrt{n} \bar{\mathcal{I}}_{\theta}(\theta_0, \hat{\pi}_n) + o_p(1)}{\bar{\mathcal{I}}_{\theta\theta}(\theta_0, \hat{\pi}_n) + \frac{1}{2}(\hat{\theta}_n - \theta_0) \bar{\mathcal{I}}_{\theta\theta\theta}(\tilde{\theta}_n, \hat{\pi}_n)} \quad (2.10)$$

We examine the numerator and denominator of the right hand side of (2.10) separately; using conditions (A1), (A2), (A3) and (A5) and the consistency of $\hat{\theta}_n$ and $\hat{\pi}_n$, we establish the following three identities:

(a) $\sqrt{n} \bar{\mathcal{I}}_{\theta}(\theta_0, \hat{\pi}_n) = \sqrt{n} \bar{\mathcal{I}}_{\theta}(\theta_0, \pi_0) - \sqrt{n}(\hat{\pi}_n - \pi_0) \mathcal{J}_{12} + o_p(1)$. To see this, we expand $\bar{\mathcal{I}}_{\theta}(\theta_0, \hat{\pi}_n)$ about π_0 , yielding

$$\begin{aligned}
\sqrt{n} \bar{\mathcal{I}}_{\theta}(\theta_0, \hat{\pi}_n) &= \sqrt{n} \bar{\mathcal{I}}_{\theta}(\theta_0, \pi_0) + \sqrt{n}(\hat{\pi}_n - \pi_0) \bar{\mathcal{I}}_{\theta\pi}(\theta_0, \pi_0) \\
&\quad + \sqrt{n}(\hat{\pi}_n - \pi_0)^2 \bar{\mathcal{I}}_{\theta\pi\pi}(\theta_0, \hat{\pi}_n) + o_p(1),
\end{aligned}$$

where $\tilde{\pi}_n$ is between π_0 and $\hat{\pi}_n$.

Since

$$\bar{\mathcal{I}}_{\theta\pi\pi}(\theta_0, \tilde{\pi}_n) = \frac{1}{n} \sum_{i=1}^n \bar{\mathcal{I}}_{\theta\pi\pi}(x_i, \theta_0, \tilde{\pi}_n),$$

we can argue as in the proof of Lemma 2.1 that $\bar{\mathcal{I}}_{\theta\pi\pi}(\theta_0, \tilde{\pi}_n)$ is bounded in probability. Since $\sqrt{n}(\hat{\pi}_n - \pi_0)^2 \xrightarrow{Pr} 0$, we have that $\sqrt{n}(\hat{\pi}_n - \pi_0)^2 \bar{\mathcal{I}}_{\theta\pi\pi}(\theta_0, \tilde{\pi}_n) = o_p(1)$. We also have

$$\sqrt{n}(\hat{\pi}_n - \pi_0) [\bar{\mathcal{I}}_{\theta\pi}(\theta_0, \pi_0) + \mathcal{J}_{12}] = o_p(1)$$

since $\sqrt{n}(\hat{\pi}_n - \pi_0)$ is bounded in probability and

$$\bar{\mathcal{I}}_{\theta\pi}(\theta_0, \pi_0) = \frac{1}{n} \sum_{i=1}^n \bar{\Phi}_{\theta\pi}(X_i, \theta_0, \pi_0)$$

converges in probability to $E_{\theta_0, \pi_0} \bar{\Phi}_{\theta\pi}(X_1, \theta_0, \pi_0) = -\mathcal{J}_{12}$. Thus, (a) is established.

(b) $\bar{\mathcal{I}}_{\theta\theta}(\theta_0, \hat{\pi}_n) \xrightarrow{\text{Pr}} -\mathcal{J}_{11}$. Define $\psi(x, \pi) = \bar{\Phi}_{\theta\theta}(x | \theta_0, \pi)$. $\psi(x, \cdot)$ is defined and differentiable for $\pi \in B$, and by (A2),

$$|\psi_\pi(x, \pi)| \leq M(x)$$

where $EM(X_1) < \infty$ and

$$E\psi(X_1, \pi_0) = -\mathcal{J}_{11}.$$

Thus, (b) follows by Lemma 2.1.

(c) $\frac{1}{2}(\hat{\theta}_n - \theta_0) \bar{\mathcal{I}}_{\theta\theta\theta}(\hat{\theta}_n, \hat{\pi}_n) \xrightarrow{\text{Pr}} 0$. To see this, it suffices to show that $\bar{\mathcal{I}}_{\theta\theta\theta}(\hat{\theta}_n, \hat{\pi}_n)$ is bounded in probability. For $\epsilon, \delta > 0$, define

$$S_n = \{(\hat{\theta}_n, \hat{\pi}_n) \in A \times B\}$$

and

$$T_n = \left\{ \frac{1}{n} \sum_{i=1}^n M(X_i) < EM(X_1) + \delta \right\}$$

where M is the dominating function in condition (A5), (i).

Since $\hat{\theta}_n \xrightarrow{\text{Pr}} \theta_0, \hat{\pi}_n \xrightarrow{\text{Pr}} \pi_0$ and $\frac{1}{n} \sum_{i=1}^n M(X_i) \xrightarrow{\text{Pr}} EM(X_1)$, we can find

$N_\epsilon \in (0, \infty)$ such that $n > N_\epsilon$ implies both $P(S_n) > 1 - \epsilon/2$ and $P(T_n) > 1 - \epsilon/2$. Thus, for $n > N_\epsilon$,

$$P\left\{\left|\frac{1}{n} \sum_{\theta\theta\theta}(\tilde{\theta}_n, \hat{\pi}_n)\right| \leq EM(X_1) + \delta\right\}$$

$$\geq P(S_n \cap T_n) > 1 - \epsilon,$$

that is, $\bar{\mathcal{I}}_{\theta\theta\theta}(\tilde{\theta}_n, \hat{\pi}_n) = O_p(1)$, establishing (c).

We thus may write

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta_0) &= \frac{-\sqrt{n} \bar{\mathcal{I}}_{\theta}(\theta_0, \hat{\pi}_n) + O_p(1)}{\bar{\mathcal{I}}_{\theta\theta}(\theta_0, \hat{\pi}_n) + \frac{1}{2}(\hat{\theta}_n - \theta_0) \bar{\mathcal{I}}_{\theta\theta\theta}(\tilde{\theta}_n, \hat{\pi}_n)} \\ &= \frac{\sqrt{n} \bar{\mathcal{I}}(\theta_0, \pi_0) - \sqrt{n}(\hat{\pi}_n - \pi_0) \mathcal{J}_{12} + O_p(1)}{\mathcal{J}_{11} + O_p(1)} \end{aligned}$$

which converges in distribution to $N(0, \sigma^2)$, where σ^2 , given in (2.9), may be obtained by noting that $\Sigma_{11} = \mathcal{J}_{11}$.

We close this section with several remarks. First, we remark that the regularity conditions under which the results of this section are proven can undoubtedly be weakened. We have made no effort to do this, but are satisfied with conditions that are fairly standard and are reasonably easy to check. Moreover, the applications of interest to us satisfy the stated conditions. We note that if $\mathcal{J}_{22} = E_{\theta_0, \pi_0} \{\dot{\Phi}_{\pi}^2\}$

exists and is positive, and if $\hat{\pi}_n$ is asymptotically equivalent to the MLE of π_0 , then the asymptotic variance of $\hat{\theta}_n$ given in (2.9) reduces to

$$\sigma^2 = \frac{J_{22}}{J_{11}J_{22} - J_{12}^2},$$

That is, $\hat{\theta}_n$ is asymptotically efficient.

III. SIGNAL PLUS NOISE MODELS

Let X be a random variable whose distribution is that of the sum of independent variables Y and Z . The distribution F_X is thus the convolution of F_Y and F_Z , and may be thought of as a model for signals in additive noise. There are many sources in nature giving rise to signal plus noise data. For example, data obtained by a Geiger counter may be viewed as sums of counts due to the presence of a radioactive substance and counts due to noise or static. Statistical communication problems in electrical engineering invariably involve signals in additive noise, usually modeled as a time series. Other examples in which convolution models appropriately describe random phenomena occur in the use of sonar for demersal fishing (see Cushing (1973), for example), in physiological processes such as synaptic transmission of neural impulses (see Katz (1966)), and in single server queue in which individuals arriving for service can be classified into mutually exclusive categories (see, for example, Shonick (1970)). Sclove and Van Ryzin (1969) describe the problem of estimating the mean density

of viruses or bacteria in a homogeneous solution where the variable observed is an area on a slide. Such a variable might well be modeled by a continuous convolution of discrete and continuous components.

Estimation problems for specific signal plus noise distributions have been examined by several authors. For example, Gaffey (1959) constructed a consistent estimator for the distribution of one component of a continuous convolution under the assumption that the "noise distribution" was known. Sclove and Van Ryzin (1969) derived method of moments estimators and their asymptotic variances for a variety of multiparameter signal plus noise distributions. Due to the cumbersome nature of the likelihood function for convolution models, maximum likelihood estimation has met substantial resistance. There has been some progress, however, in one parameter problems. For example, Samaniego (1976) proved the following result:

Theorem 3.1. Let X be a nonnegative integer valued random variable whose distribution is indexed by a positive parameter θ . Suppose the probability mass function is differentiable in θ . Then

$$\frac{\partial}{\partial \theta} P_{\theta}(X=n) = P_{\theta}(X=n-1) - P_{\theta}(X=n) \quad \forall n, \forall \theta \quad (3.1)$$

if, and only if, the distribution of X is a convolution of the Poisson distribution with parameter θ and the distribution of a nonnegative integer valued random variable which is independent of θ .

If X_1, \dots, X_n is a sample from a convoluted Poisson distribution, Theorem 3.1 implies that the likelihood equation $\mathcal{L}_\theta(\theta) = 0$ may be written as

$$\sum_{i=1}^n \frac{P_\theta(X_i = x_i - 1)}{P_\theta(X_i = x_i)} = n,$$

where \tilde{x} is the vector of observations. It is thus clear that the behavior of probability ratios such as

$$\frac{P_\theta(X=a)}{P_\theta(X=b)} \tag{3.2}$$

for $a < b$ is relevant to this estimation problem. The property parametric monotone decreasing ratio (PMDR) is defined in Samaniego (1976) to mean that ratios of the form (3.2) are decreasing. For convoluted Poisson distributions with PMDR, the likelihood equation has at most one solution, and the MLE is easily found numerically. It is shown that while there are convoluted Poisson distributions for which the likelihood equation may have any fixed number of solutions, there are many convoluted Poisson distributions which indeed have the PMDR property. Among these is the Poisson-binomial convolution, which we denote by $P(\theta) * B(N, p)$, with p known. This latter model, with both parameters unknown, is one of several models examined in the next section. Similar results to those above are obtained in Samaniego (1977) for known convolutions of binomial distributions. Convoluted Pascal distributions are studied in Samaniego and Hannon (1978).

One-parameter problems such as those described above provide the key to the feasibility of pseudo maximum likelihood estimation in convolution models. Not only is it true that in many situations in which a convolution model is appropriate, there is a parameter of particular interest, but it is also true for many convolution models that when all parameters save the one of interest are replaced by estimates, the pseudo MLE of the remaining parameter is easily found. The convolution $\mathcal{P}(\theta) * \mathcal{B}(N, p)$ is a good example of a problem in which maximum likelihood estimation of (θ, p) is quite difficult, but pseudo maximum likelihood estimation of either parameter is easily accomplished.

In the remainder of this section, we develop some new results concerning continuous convolution models. Specifically, we will consider two convolutions of normal distributions, and demonstrate that they share the characteristics in one-parameter estimation problems that have been established for the discrete models discussed above. We adapt the following definition from Samaniego (1977).

Definition 3.1. Let $\{f_{\theta, \alpha}, \theta \in \Theta, \alpha \in \mathcal{A}\}$ be a family of density functions indexed by a real parameter θ and a (possibly degenerate) parameter α . The family is said to have parametric monotone decreasing ratio (PMDR) in θ if, for each fixed α and for each $x < u$ for which $f(\cdot | p, \alpha) > 0$ for every $\theta \in \Theta, \alpha \in \mathcal{A}$, the ratio

$$\frac{f(x | \theta, \alpha)}{f(u | \theta, \alpha)}$$

is decreasing in θ .

Let X be a continuous random variable distributed as the sum of a binomial variable $Y \sim B(N, p)$ and a continuous variable Z whose distribution does not depend on p . Let $g(\cdot)$ be the density of Z . Then the density of X is given by

$$f(x|N, p) = \sum_{y=0}^N g(x-y) \binom{N}{y} p^y (1-p)^{N-y} \quad (3.1)$$

It is easy to verify that this density satisfies the system of differential equations

$$\frac{\partial}{\partial p} f(x|N, p) = N[f(x-1|N-1, p) - f(x|N-1, p)] \quad (3.2)$$

for all x and for all $p \in (0, 1)$. It follows, given a sample of size n , that the likelihood equation $\mathfrak{L}_p(p) = 0$ is given by

$$\sum_{i=0}^n \frac{f(x_i-1|N-1, p) - f(x_i|N-1, p)}{f(x_i|N, p)} = 0, \quad (3.3)$$

where \tilde{x} is the vector of observations. Rewriting the summand in (3.3) as

$$\frac{f(x_i-1|N-1, p) - f(x_i|N-1, p)}{p[f(x_i-1|N-1, p) - f(x_i|N-1, p)] + f(x_i|N-1, p)}$$

makes it easy to argue that if a continuous convolution of the binomial distribution has PMDR, then the logarithmic derivative $\mathfrak{L}_p(p)$ of the likelihood is decreasing in p . Thus, the MLE for p in a one-parameter model of the form (3.1) with PMDR is either zero, one of the unique solution of the likelihood equation (3.3).

The development of the preceeding paragraph holds as well for continuous convolutions of the Poisson distribution. Specifically, let X be a random variable distributed as the sum of a Poisson variable $X \sim P(\theta)$ and a continuous variable Z whose distribution does not depend on θ . If Z has density g , then the density of X is given by

$$f(x|\theta) = \sum_{y=0}^{\infty} g(x-y) \frac{\theta^y e^{-\theta}}{y!} \quad (3.4)$$

Densities of the form (3.4) satisfy the system of differential equations

$$\frac{\partial}{\partial \theta} f(x|\theta) = f(x-1|\theta) - f(x|\theta) \quad (3.5)$$

for all x and for all θ . It follows that the likelihood equation for a sample of size n from this distribution is given by

$$\sum_{i=1}^n \frac{f(x_i-1|\theta)}{f(x_i|\theta)} = n. \quad (3.6)$$

It is clear that when a continuous convolution of the Poisson distribution has PMDR, the likelihood equation has at most one root, and the MLE may be easily obtained numerically.

We now demonstrate that several important continuous convolution models indeed have PMDR in the parameter of the discrete component. We consider below normal convolutions of binomial and Poisson distributions. Since the parameters of the normal are assumed known, we may, without loss of generality, set $\mu = 0$ and $\sigma^2 = 1$.

Theorem 3.2. The convolution $\beta(N,p) * h(0,1)$ has PMDR in p .

Proof. We denote the normal density by $n(\cdot)$. Let $x < u$, and consider the ratio

$$\begin{aligned} R(p;x,u) &= \frac{f(x|p)}{f(u|p)} \\ &= \frac{\sum_{y=0}^N n(x-y) \binom{N}{y} p^y (1-p)^{N-y}}{\sum_{v=0}^N n(u-v) \binom{N}{v} p^v (1-p)^{N-v}} \\ &= \frac{\sum_{y=0}^N n(x-y) \binom{N}{y} \left(\frac{p}{1-p}\right)^y}{\sum_{v=0}^N n(u-v) \binom{N}{v} \left(\frac{p}{1-p}\right)^v} \end{aligned}$$

We will show $R(p;x,u)$ is decreasing in p . To this end, we define the function

$$g_{x,y}(\theta) = \frac{\sum_{y=0}^N n(x-y) \binom{N}{y} \theta^y}{\sum_{v=0}^N n(u-v) \binom{N}{v} \theta^v}.$$

We intend to show that $\frac{\partial}{\partial \theta} g_{x,y}(\theta) < 0$. We note that the numerator of

$\frac{\partial}{\partial \theta} g_{x,y}(\theta)$ is

$$\begin{aligned}
& \left(\sum_{y=1}^N n(x-y) \binom{N}{y} y \theta^{y-1} \right) \left(\sum_{v=0}^N n(u-v) \binom{N}{v} \theta^v \right) \\
& - \left(\sum_{y=0}^N n(x-y) \binom{N}{y} \theta^y \right) \left(\sum_{v=1}^N n(u-v) \binom{N}{v} v \theta^{v-1} \right) \\
& = n(u) \sum_{y=1}^N n(x-y) \binom{N}{y} y \theta^{y-1} \\
& + n(x) \sum_{v=1}^N n(u-v) \binom{N}{v} v \theta^{v-1} \\
& + \left(\sum_{y=1}^N n(x-y) \binom{N}{y} y \theta^{y-1} \right) \left(\sum_{v=1}^N n(u-v) \binom{N}{v} \theta^v \right) \\
& + \left(\sum_{y=1}^N n(x-y) \binom{N}{y} \theta^y \right) \left(\sum_{v=1}^N n(u-v) \binom{N}{v} v \theta^{v-1} \right) \\
& = A + B
\end{aligned}$$

where

$$A = \sum_{y=1}^N [n(x-y)n(u) - n(u-y)n(x)] \binom{N}{y} y \theta^{y-1}$$

and

$$B = \sum_{y=1}^N \sum_{v=1}^N n(x-y)n(u-v) \binom{N}{y} \binom{N}{v} (y-v) \theta^{y+v-1}.$$

Now $A < 0$, since the inequality

$$n(x-y)n(u) - n(u-y)n(x) < 0$$

is easily shown to be equivalent to

$$x < u.$$

To show that $B < 0$, we define, for fixed $x < u$, the function

$$G(y,v) = n(x-y)n(u-v) \binom{N}{y} \binom{N}{v}.$$

One may show that the inequality

$$G(y,v) > G(v,y)$$

is equivalent to the inequality

$$y < v.$$

Thus, we may write

$$\begin{aligned} B &= \sum_{y=1}^N \sum_{v=1}^N G(y,v) (y-v) e^{y+v-1} \\ &= \sum_{y < v} G(y,v) (y-v) e^{y+v-1} + \sum_{y > v} G(y,v) (y-v) e^{y+v-1} \\ &= \sum_{y < v} G(y,v) (y-v) e^{y+v-1} + \sum_{y < v} G(v,y) (v-y) e^{y+v-1} \\ &= \sum_{y < v} [G(y,v) - G(v,y)] (y-v) e^{y+v-1}. \end{aligned}$$

This latter sum is clearly negative. We thus have that $g_{x,y}(\theta)$ is a decreasing function of θ . Since

$$R(p;x,u) = g_{x,u} \cdot \theta(p)$$

where $g_{x,u}$ is decreasing and θ is increasing, we have that $R(\cdot;x,u)$ is decreasing, completing the proof.

Theorem 3.3. The convolution $\mathcal{P}(\theta) * n(0,1)$ has PMDR in θ .

Proof. This result obtains using the same style of proof as in Theorem 3.2.

PMDR may be demonstrated for a large class of continuous convolutions of binomial or Poisson distributions. For example, it is easy to show that the convolution $\mathcal{P}(\theta) * U[a,b]$ of Poisson and uniform distributions has PMDR in θ . Our original proof of Theorem 3.3 was based on a limiting argument which approximated the distribution of $Y + Z$, with $Z \sim n(0,1)$ by $Y + \frac{1}{\sqrt{k}} \sum_{j=1}^k Z_j$ where $Z_j \stackrel{iid}{\sim} U[-\sqrt{3}, \sqrt{3}]$. We showed that each approximating distribution had PMDR in θ and argued that the limiting distribution $Y + Z$ must also have PMDR in θ . This method of proof, which is quite inefficient by comparison to the direct proofs of Theorems 3.2 and 3.3, does suggest that the class of continuous convolutions of Poisson or binomial distributions having PMDR is rather broad.

We remark in closing this section that the system of differential equations in (3.2) and (3.5) have not been shown to characterize continuous convolutions of binomial or Poisson distributions. We conjecture that they do.

IV. REGULARITY OF SIGNAL PLUS NOISE MODELS

In this section, we seek to demonstrate that a variety of signal plus noise models satisfy the regularity conditions under which the asymptotic theory of pseudo maximum likelihood estimates has been developed. Since demonstrations of this sort are tedious and uninteresting, we give details only for one fairly typical model - the convolution of a Poisson signal distribution with a normal noise distribution. We state without proof that three other models are also regular: Poisson signals in binomial noise, binomial signals in Poisson noise and binomial signals in normal noise. The verification of regularity conditions for these three models is no more complex than that for the model considered here. The regularity of these models, together with the fact that method of moments estimators of noise parameters satisfy the requirements of Theorems 2.1 and 2.2, renders as fully known the asymptotic behavior of at least one pseudo maximum likelihood estimator of the signal parameter of each model.

Let X be a random variable distributed as the sum $Y + Z$ of independent variables, where Y is Poisson distributed with parameter θ and Z is normally distributed with mean zero and variance σ^2 . The distribution of X will be denoted by $P(\theta) * N(0, \tau)$, where $\tau = \sigma^{-2}$;

the mass function of Y is denoted by $P(y|\theta)$, $y=0,1,\dots$ and the density of Z is denoted by $n(z|\tau)$. The density of X is given by

$$f(x|\theta, \tau) = \sum_{y=0}^{\infty} n(x-y|\tau)P(y|\theta), \quad -\infty < x < \infty. \quad (4.1)$$

Let (θ_0, τ_0) be the true parameter vector, and assume $0 < a_1 < \theta_0 < a_2 < \infty$ and $0 < b_1 < \tau_0 < b_2 < \infty$. The open sets A and B introduced in Section II represent the open intervals (a_1, a_2) and (b_1, b_2) respectively. We first prove a theorem which will be useful in the verification of several regularity conditions. We need the following result.

Lemma 4.1. Let $f(x|\theta, \tau)$ be a density of the form (4.1). Then

$$\frac{\partial}{\partial \tau} f(x|\theta, \tau) = \frac{1}{2} \left[\left(\frac{1}{\tau} - x^2 \right) f(x|\theta, \tau) - \theta(1-2x)f(x-1|\theta, \tau) - \theta^2 f(x-2|\theta, \tau) \right] \quad (4.2)$$

Proof. First note that

$$\frac{\partial}{\partial \tau} n(x|\tau) = \frac{1}{2} \left(\frac{1}{\tau} - x^2 \right) n(x|\tau).$$

In differentiating $f(x|\theta, \tau)$, we need to pass a derivative through an infinite sum, a valid operation since

$$\begin{aligned} & \left| \frac{\partial}{\partial \tau} n(x-y|\tau)P(y|\theta) \right| \\ &= \frac{1}{2} \left| \frac{1}{\tau} - (x-y)^2 \right| n(x-y|\tau)P(y|\theta) \\ &\leq \frac{1}{2} \left(\frac{1}{b_2} + (x-y)^2 \right) n(0|b_2)P(y|\theta) \\ &\equiv M(y, \theta). \end{aligned}$$

and, for each fixed $\theta \in A$, $\sum_{y=0}^{\infty} M(y, \theta) < \infty$. Thus

$$\begin{aligned} \frac{\partial}{\partial \tau} f(x|\theta, \tau) &= \sum_{y=0}^{\infty} \frac{\partial}{\partial \tau} n(x-y|\tau) p(y|\theta) \\ &= \sum_{y=0}^{\infty} \frac{1}{2} \left(\frac{1}{\tau} - (x-y)^2 \right) n(x-y|\tau) p(y|\theta). \end{aligned} \quad (4.3)$$

To establish the desired identity, we deal with the probability mass function

$$g(y) = \frac{n(x-y|\tau) p(y|\theta)}{f(x|\theta, \tau)} \quad y=0, 1, \dots$$

where θ, τ and x are considered fixed. One can easily verify that if Y is a nonnegative integer valued random variable with mass function $g(y)$, then

$$EY = \theta \frac{f(x-1|\theta, \tau)}{f(x|\theta, \tau)}$$

and
$$EY(Y-1) = \theta^2 \frac{f(x-2|\theta, \tau)}{f(x|\theta, \tau)},$$

from which it follows that

$$EY^2 = \theta^2 \frac{f(x-2|\theta, \tau)}{f(x|\theta, \tau)} + \theta \frac{f(x-1|\theta, \tau)}{f(x|\theta, \tau)}$$

We may thus expand, for fixed x ,

$$\begin{aligned}\Sigma(x-y)^2 n(x-y|\tau) p(y|\theta) &= E(x-Y)^2 f(x|\theta, \tau) \\ &= x^2 f(x|\theta, \tau) + \theta(1-2x)f(x-1|\theta, \tau) + \theta^2 f(x-2|\theta, \tau).\end{aligned}$$

Substitution of the latter expression into (4.3) yields the identity (4.2).

Theorem 4.1. Let $\{f(x|\theta, \tau)\}$ be the class of density functions on the real line of the form (4.1), and let C be the class of all linear combinations of terms of the form

$$\alpha(x, \theta, \tau) \prod_{i=1}^m \frac{f(a_i|\theta, \tau)}{f(x|\theta, \tau)}, \quad (4.4)$$

where $\alpha(x, \theta, \tau)$ is a polynomial in x , θ and $1/\tau$, and a_i , $i=1, \dots, m$ are real numbers such that $x - a_i$ is a positive integer for each i . Then

- (i) The class C is closed under partial differentiation with respect to θ or τ , and
- (ii) Every member of C is bounded on $A \times B$ by a function with finite expectation.

Proof. To prove (i), we need only examine a single term of the form (4.4). Suppose $\alpha(x, \theta, \tau) = \sum_{j,k,l} \beta_{jkl} \frac{x^j \theta^k}{\tau^l}$, and denote $f(\cdot|\theta, \tau)$ by $f(\cdot)$ for simplicity. Then

$$\begin{aligned}
& \frac{\partial}{\partial \theta} \alpha(x, \theta, \tau) \prod_{i=1}^m \frac{f(a_i)}{f(x)} \\
&= \left[\sum_k \beta_{jkl} \frac{x^j \theta^{k-1}}{\tau^l} \right] \prod_{i=1}^m \frac{f(a_i)}{f(x)} \\
&+ \alpha(x, \theta, \tau) \frac{\sum_{i=1}^m [f(a_i-1) - f(a_i)] \prod_{r \neq i} f(a_r)}{(f(x))^m} \\
&+ \alpha(x, \theta, \tau) \prod_{i=1}^m f(a_i) \cdot \frac{1}{f(x)^{m+1}} \cdot (-m) [f(x-1) - f(x)],
\end{aligned}$$

which, by inspection, is in the class C. We also have, using Lemma 4.1, that

$$\begin{aligned}
& \frac{\partial}{\partial \tau} \alpha(x, \theta, \tau) \prod_{i=1}^m \frac{f(a_i)}{f(x)} = \left[\sum (-l) \beta_{jkl} \frac{x^j \theta^k}{\tau^{l+1}} \right] \prod_{i=1}^m \frac{f(a_i)}{f(x)} \\
&+ \frac{\alpha(x, \theta, \tau)}{2(f(x))^m} \cdot \sum_{i=1}^m \left[\left(\frac{1}{\tau} - a_i \right)^2 f(a_i) - \theta(1-2a_i)f(a_i-1) \right. \\
&\quad \left. - \theta^2 f(a_i-2) \right] \prod_{r \neq i} f(a_r) \\
&+ \alpha(x, \theta, \tau) \prod_{i=1}^m f(a_i) \cdot \frac{1}{2(f(x))^{m+1}} \cdot (-m) \left[\left(\frac{1}{\tau} - x^2 \right) f(x) - \theta(1-2x)f(x-1) \right. \\
&\quad \left. - \theta^2 f(x-2) \right]
\end{aligned}$$

is in the class C. This completes the proof of (i). To prove (ii), we show that a single term of the form (4.4) is bounded on $A \times B$ by a

function with finite expectation. If $\alpha(x, \theta, \tau) = \sum_{j,k,l} \beta_{jkl} \frac{x^j \theta^k}{\tau^l}$, then

$$|\alpha(x, \theta, \tau)| \leq \sum |\beta_{jkl}| \frac{|x|^j \theta^k}{\tau^l}$$

Since the convolution $P(\theta) * h(0, \tau)$ has PMDR in θ , we have for any integer i

$$\begin{aligned} \frac{f(x-i|\theta, \tau)}{f(x|\theta, \tau)} &\leq \frac{f(x-i|0, \tau)}{f(x|0, \tau)} \\ &= \frac{n(x-i|\tau)}{n(x|\tau)} \\ &= e^{\tau i x} e^{-\tau i^2/2} \\ &\leq \left[e^{b_1 i x} + e^{b_2 i x} \right] e^{-b_1 i^2/2} \end{aligned}$$

Thus, terms of the form (4.4) can be bounded above by linear combinations of terms like $|x|^i e^{b x}$ for specific positive integers i and positive constants b . It remains to show that for any positive integer i and any $b > 0$, $E|X|^i e^{b X} < \infty$. First we note that for any positive i and b ,

$$\sum_{y=0}^{\infty} y^i e^{by} P(y|\theta) = e^{\theta(1-e^b)} \sum y^i P(y|\theta e^b) < \infty \quad (4.5)$$

and

$$\int_{-\infty}^{\infty} |x|^i e^{b x} n(x|\tau) dx = e^{b^2/2\tau} \int_{-\infty}^{\infty} |x|^i \sqrt{\frac{\tau}{2\pi}} e^{-(\tau/2)(x-(b/2))^2} dx < \infty \quad (4.6)$$

Now, for $X \sim P(\theta) * n(0, \tau)$, we have, using Fubini's theorem and an elementary inequality,

$$\begin{aligned}
 E|X|^1 e^{bX} &= \int_{-\infty}^{\infty} |x|^1 e^{bx} f(x|\theta, \tau) dx \\
 &= \int_{-\infty}^{\infty} |x|^1 e^{bx} \sum_{y=0}^{\infty} n(x-y|\tau) P(y|\theta) dy \\
 &= \sum_{y=0}^{\infty} P(y|\theta) \int_{-\infty}^{\infty} |x|^1 e^{bx} n(x-y|\tau) dx \\
 &= \sum_{y=0}^{\infty} P(y|\theta) e^{by} \int_{-\infty}^{\infty} |(u+y)|^1 e^{bu} n(u|\tau) du
 \end{aligned}$$

(where $u = x-y$)

$$\begin{aligned}
 &\leq \sum_{y=0}^{\infty} P(y|\theta) e^{by} \int_{-\infty}^{\infty} 2^1 (|u|^1 + |y|^1) e^{bu} n(u|\tau) du \\
 &= 2^1 \sum_{y=0}^{\infty} P(y|\theta) e^{by} \int_{-\infty}^{\infty} |u|^1 e^{bu} n(u|\tau) du \\
 &\quad + 2^1 \sum_{y=0}^{\infty} |y|^1 e^{by} P(y|\theta) \int_{-\infty}^{\infty} e^{bu} n(u|\tau) du.
 \end{aligned}$$

The latter terms are finite by (4.5) and (4.6).

We now demonstrate the regularity of the model $P(\theta) * h(0, \tau)$.

(A1) Derivatives of Φ of all orders in θ and τ exist and may be written in closed form as functions of the density $f(\cdot|\theta, \tau)$ in (4.1). This claim follows from part (i) of Theorem 4.1 and the fact that the derivatives of Φ_θ and Φ_τ belong to the class C of Theorem 4.1. Indeed,

$$\Phi_\theta(x|\theta, \tau) = \frac{f(x-1|\theta, \tau) - f(x|\theta, \tau)}{f(x|\theta, \tau)}$$

and

$$\Phi_\tau(x|\theta, \tau) = \frac{(\frac{1}{\tau} - x^2)f(x|\theta, \tau) - \theta(1-2x)f(x-1|\theta, \tau) - \theta^2 f(x-2|\theta, \tau)}{2f(x|\theta, \tau)}$$

(A2) The validity of the interchange of differentiation and integration of the density f may be checked directly. For example, for any $\theta > 0$ and $\tau > 0$,

$$\frac{\partial}{\partial \theta} \int_R f(x|\theta, \tau) du = 0$$

and

$$\int_R \frac{\partial}{\partial \theta} f(x|\theta, \tau) du$$

$$\int_R (f(x-1|\theta, \tau) - f(x|\theta, \tau)) du = 0$$

Other interchanges may be checked similarly.

(A3) The existence of \mathcal{J}_{11} and \mathcal{J}_{12} follow from Theorem 4.1. That $\mathcal{J}_{11} > 0$ is obvious.

(A4) We have

$$\begin{aligned} & \left| \frac{\partial}{\partial \tau} \log \frac{f(x|\theta, \tau)}{f(x|\theta_0, \tau)} \right| \\ &= \left| \frac{\partial}{\partial \tau} \log f(x|\theta, \tau) - \frac{\partial}{\partial \tau} \log f(x|\theta_0, \tau) \right| \\ &\leq |\dot{\Phi}_\tau(x|\theta, \tau)| + |\dot{\Phi}_\tau(x|\theta_0, \tau)| \end{aligned}$$

The latter two terms are bounded on $A \times B$ by functions with finite expectation according to Theorem 4.1.

(A5) Partial derivatives of $\dot{\Phi}$ of any order are members of the class C and are thus bounded on $A \times B$ by integrable functions.

(A6) The identifiability of the model $\mathcal{P}(\theta) * \mathcal{N}(0, \tau)$ is an easy consequence of Theorem 2 of Sclove and Van Ryzin (1969). Using their notation, we may define

$$H_1(X_1, \dots, X_n) = \bar{X}_n$$

and

$$H_2(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = S^2.$$

Then

$$h_1(\theta, \tau) = EH_1(X_1, \dots, X_n) = \theta$$

$$h_2(\theta, \tau) = EH_2(X_1, \dots, X_n) = \theta + \frac{1}{\tau}.$$

Since θ and $\frac{1}{\tau}$ are strictly monotone functions, we have from the aforementioned theorem that $\mathcal{P}(\theta) * \mathcal{N}(0, \tau)$ is identifiable.

V. EFFICIENCY QUESTIONS

For the signal plus noise models discussed in the previous section, we may now claim the consistency and asymptotic normality of the pseudo maximum likelihood estimate of the signal parameter, provided the noise parameter is estimated appropriately. The method of moments, for example, yields estimates of the noise parameter that satisfy the requirements of Theorems 2.1 and 2.2. The efficiency of the resulting pseudo maximum likelihood estimates will now be examined. We present below evidence in support of our conjecture that these pseudo MLE's are uniformly more efficient asymptotically than the method of moments estimators of the signal parameter. Our results suggest that solving a reduced system of likelihood equations for certain parameters of a model serves to improve the asymptotic efficiency of estimates of these parameters. We are unable at present to prove our conjecture due to the fact that the asymptotic variances to be compared are functions of the Fisher information matrix, the components of which are only available via series representations. However, since we are able to approximate these asymptotic variances to any desired degree of accuracy for any fixed value of the parameter pair, we have computed and compared them for the Poisson-Binomial convolution for a fairly broad collection of parameter values. We study the asymptotic relative efficiency of PMLE's both for Poisson signals in Binomial noise and for Binomial signals in Poisson noise.

with

$$t = \frac{1}{2Np_0}$$

$$\Gamma_{22} = \theta_0 + Np_0(1-p_0)$$

$$\Gamma_{23} = \theta_0 + Np_0(1-p_0)(1-2p_0)$$

and
$$\Gamma_{23} = \theta_0 + 2[\theta_0 + Np_0(1-p_0)]^2 + Np_0(1-p_0)[1-6p_0(1-p_0)]^2$$

To see this, we note that the vector

$$U = \sqrt{n} \begin{bmatrix} \bar{x}_\theta(\theta_0, p_0) \\ H_1 - \mu \\ H_2 - \sigma^2 \end{bmatrix},$$

where $\mu = EX_1$ and $\sigma^2 = E(X_1 - \mu)^2$, is asymptotically equivalent to the vector

$$\begin{bmatrix} \frac{1}{n} \sum_{i=1}^n \bar{\Phi}_\theta(X_i | \theta_0, p_0) \\ \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \\ \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \sigma^2 \end{bmatrix}.$$

It follows from the multivariate central limit theorem that

$$U \xrightarrow{D} n(O, \Gamma),$$

where

$$\Gamma_{11} = \text{var } \bar{\Phi}_{\theta}(X_1 | \theta_0, p_0) = \mathcal{J}_{11}$$

$$\Gamma_{12} = \text{cov}(\bar{\Phi}_{\theta}(X_1 | \theta_0, p_0), X_1)$$

$$= E(X_1 \bar{\Phi}_{\theta}(X_1 | \theta_0, p_0)) \quad \text{by condition (A2)}$$

$$= \sum_{x=0}^{\infty} x \frac{\partial}{\partial \theta} f(x | \theta, p_0) \Big|_{\theta_0}$$

$$= \sum_{x=0}^{\infty} x [f(x-1 | \theta_0, p_0) - f(x | \theta_0, p_0)]$$

$$= 1$$

$$\Gamma_{13} = \text{cov}(\bar{\Phi}_{\theta}(X_1 | \theta_0, p_0), (X_1 - \mu)^2)$$

$$= E(X_1 - \mu)^2 \bar{\Phi}_{\theta}(X_1 | \theta_0, p_0)$$

$$= \sum_{x=0}^{\infty} (x - \mu)^2 [f(x-1 | \theta_0, p_0) - f(x | \theta_0, p_0)]$$

$$= 1$$

$$\Gamma_{22} = \text{var } X_1 = \theta_0 + N p_0 (1 - p_0)$$

$$\Gamma_{23} = \text{cov}(X_1 - \mu, (X_1 - \mu)^2) = \theta_0 + N p_0 (1 - p_0) (1 - 2p_0)$$

$$\begin{aligned} \text{and } \Gamma_{33} = \text{var}(X_1 - \mu)^2 &= \theta_0 + 2[\theta_0 + N p_0 (1 - p_0)]^2 \\ &+ N p_0 (1 - p_0) [1 - 6p_0 (1 - p_0)] . \end{aligned}$$

We may obtain (5.1) by the δ -method, since

$$\begin{bmatrix} \bar{\mathcal{J}}_{\theta}(\theta_0, p_0) \\ \tilde{\pi}_n \end{bmatrix} = g \begin{bmatrix} \mathcal{J}_{\theta}(\theta_0, p_0) \\ H_1 \\ H_2 \end{bmatrix},$$

where

$$g_1(t_1, t_2, t_3) = t_1$$

and

$$g_2(t_1, t_2, t_3) = \sqrt{\frac{t_2 - t_3}{N}},$$

is a totally differentiable transformation.

Let

$$A = \begin{vmatrix} 1 & 0 & 0 \\ 0 & t & -t \end{vmatrix}.$$

Then (5.1) follows from the δ -method theorem (see Rao (1965)), with

$$\Sigma = A \Gamma A^T.$$

We now record the asymptotic variances of the maximum likelihood estimate (MLE), the pseudo maximum likelihood estimate (PMLE) and the method of moments estimate (MME) of the signal parameter θ_0 :

$$\sigma_{MLE}^2 = \frac{\mathcal{J}_{22}}{\mathcal{J}_{11}\mathcal{J}_{22} - \mathcal{J}_{12}^2} \quad (5.2)$$

$$\sigma_{PMLE}^2 = \frac{1}{\mathcal{J}_{11}} + \frac{\mathcal{J}_{12}^2}{\mathcal{J}_{11}^2} t^2 [\Gamma_{22} - 2\Gamma_{23} + \Gamma_{33}] \quad (5.3)$$

$$\sigma_{MME}^2 = (1 - Nt)^2 \Gamma_{22} + 2Nt(1 - Nt)\Gamma_{23} + (Nt)^2 \Gamma_{33}. \quad (5.4)$$

The difficulty in directly comparing these expressions resides in our inability to represent \mathcal{J}_{ij} in any form other than the series below:

$$\begin{aligned} \mathcal{J}_{11} &= E \mathcal{J}_{\theta}^2(X|\theta_0, p_0) \\ &= \sum_{x=0}^{\infty} \frac{[f(x-1|N, \theta_0, p_0) - f(x|N, \theta_0, p_0)]^2}{f(x|N, \theta_0, p_0)} \end{aligned} \quad (5.5)$$

$$\begin{aligned}
J_{12} &= E \bar{\Phi}_{\theta}(X|\theta_0, p_0) \bar{\Phi}_p(X|\theta_0, p_0) \\
&= N \sum_{x=0}^{\infty} \frac{[f(x-1|N, \theta_0, p_0) - f(x|N, \theta_0, p_0)][f(x-1|N-1, \theta_0, p_0) - f(x|N-1, \theta_0, p_0)]}{f(x|N, \theta_0, p_0)}
\end{aligned}
\tag{5.6}$$

$$\begin{aligned}
J_{22} &= E \bar{\Phi}_p^2(X|\theta_0, p_0) \\
&= N^2 \sum_{x=0}^{\infty} \frac{[f(x-1|N-1, \theta_0, p_0) - f(x|N-1, \theta_0, p_0)]^2}{f(x|N, \theta_0, p_0)}
\end{aligned}
\tag{5.7}$$

We approximate J_{ij} by sums of the initial terms of the series (5.5)-(5.7). We develop bounds on the error of approximating the infinite series (5.5)-(5.7) by finite sums. We give details only for the approximation of (5.5). First, note that the ratio

$$g(x) = \frac{f(x-1|N, \theta, p)}{f(x|N, \theta, p)}$$

is decreasing in p (this is the PMDR property). Since $\bar{\Phi}_{\theta}(x) = q(x)-1$, we have

$$|\bar{\Phi}_{\theta}(x)| \leq \max \{ |\bar{\Phi}_{\theta}(x|N, \theta, 0)|, |\bar{\Phi}_{\theta}(x|N, \theta, 1)| \}.$$

From this it follows that

$$|\bar{\Phi}_{\theta}(x)| \leq \frac{x-\theta}{\theta},$$

provided $x \geq L \geq N+\theta$. We thus have for $L > N+\theta$, the error

$$\begin{aligned}
E_{11} &\equiv J_{11} - \sum_{x=0}^L \bar{\phi}_{\theta}^2(x|N, \theta_0, p_0) f(x|N, \theta_0, p_0) \\
&= \sum_{x=L+1}^{\infty} \bar{\phi}_{\theta}^2(x) f(x) \\
&\leq \sum_{x=L+1}^{\infty} \frac{(x-\theta)^2}{\theta^2} f(x) \\
&= E \frac{(x-\theta)^2}{\theta^2} - \sum_{x=0}^L \frac{(x-\theta)^2}{\theta^2} f(x).
\end{aligned}$$

Similar bounds may be established for

$$E_{12} = J_{12} - \sum_{x=0}^L \bar{\phi}_{\theta}(x) \bar{\phi}_p(x)$$

and

$$E_{22} = J_{22} - \sum_{x=0}^L \bar{\phi}_p^2(x).$$

With such bounds, we may approximate $J_{ij}(\theta_0, p_0)$ to any desired accuracy for any fixed value of the pair (θ_0, p_0) . For a specific range of parameters, we have approximated J_{ij} with errors E_{ij} no greater than 10^{-7} , and have used these approximations in computing the asymptotic efficiency of the PMLE relative to the MLE and of the MME relative to the PMLE. The parameter values examined are the following:

$$\begin{aligned}
\theta_0 &= .1, .25, .5, 1(1)10 \\
p_0 &= .1(.1).9 \\
N &= 1(1)10.
\end{aligned} \tag{5.8}$$

For the parameter values above, we find

$$.57288 \leq \text{ARE}(\text{PMLE}/\text{MLE}) < 1$$

$$.42712 \leq \text{ARE}(\text{MME}/\text{PMLE}) < 1.$$

From this computation, we are led to conjecture that this pseudo maximum likelihood estimate of θ is uniformly more efficient asymptotically than the moment estimator of θ .

We obtained similar results for the pseudo MLE of the binomial parameter using the moment estimator of the Poisson parameter. Specifically, for the parameter set in (5.8), we find

$$.63983 \leq \text{ARE}(\text{PMLE}/\text{MLE}) < 1$$

$$.34036 \leq \text{ARE}(\text{MME}/\text{PMLE}) < 1.$$

VI. DISCUSSION

This paper advances a method of estimating a subset of parameters in multiparameter models, and investigates its numerical and asymptotic characteristics. We have given general conditions under which pseudo maximum likelihood estimates are consistent and asymptotically normal. For the signal plus noise models discussed in Section IV, we have demonstrated the numerical feasibility of the approach and have verified that the asymptotic properties of the general method obtain. Moreover, a numerical investigation suggests that pseudo maximum likelihood estimates lie strictly between the MLE and the method of moments estimate for the signal parameter in terms of asymptotic efficiency. In general, we view the process of solving a reduced system of likelihood equations as a technique which promises to improve the asymptotic

behavior of estimates of specific parameters.

The models to which we have applied the method of pseudo maximum likelihood estimation constitute a large class of models of considerable practical interest. We do not wish to leave the impression, however, that the method is limited to this particular application. The method might be considered for estimating parameters of any model for which consistent estimates are readily available, while the construction of efficient estimates poses an untractable problem. The method of moments has been shown by a number of authors to be tractable for many mixture models for which direct maximum likelihood estimation is virtually impossible. There are also a number of models popularly used in reliability studies for which lower dimensional maximum likelihood estimation is straightforward and method of moments estimators as well as other reasonable but suboptimal estimators are easily obtained for the full set of parameters. Among these are three parameter gamma distributions, two or three parameter Weibull distributions, and the extreme value distributions. Details on estimation for these models are nicely summarized in Johnson and Kotz (1970). Another potential application of the pseudo maximum likelihood approach is to the estimation of regression parameters of variables transformed to normality via the Box-Cox (1964) transformation. Finally, there are many nuisance parameter models that have been studied or mentioned in the literature (see, for example, Basu (1975) and (1977)), and estimation for some of these models may well benefit from application of the approach we have advanced here. Indeed, we believe that this paper will provide

theoretical justification for what practitioners have been doing for years.

It is worthwhile calling attention to the fact that the asymptotic theory developed here is not directly comparable to the asymptotic theory for maximum likelihood estimates or, more generally, for best asymptotically normal (BAN) estimates. In a sense, the requirements on the model are slightly less stringent for pseudo maximum likelihood estimation than for these other methods. Specifically, the asymptotic results obtained in Section II require a little less regularity in the model vis-a-vis the nuisance parameter than is required by the asymptotic theory of efficient estimates. Immediate evidence of this is the fact that the Fisher information

$$J_{22} = E(\dot{\phi}_p^2(X|\theta, p))$$

need not exist, and does not enter into the results proven here. There are other differences in regularity conditions, but a full discussion of them is somewhat academic. The models we have examined are fully regular under either theory, as are most models of interest. It remains true that the method of pseudo maximum likelihood estimation may be applicable in situations where the standard theory for MLE's or BAN estimates breaks down. We leave this fact for the interested reader to explore further.

We make one final remark on BAN estimation. A common approach to BAN estimation is the so-called expansion of a \sqrt{n} -consistent estimate in a Taylor series, which may also be viewed as a one-step iteration using the Newton-Raphson procedure. Since the signal plus noise models

we have examined are BAN-regular, the method may be used with the method of moments estimator as the seed. The limitation one encounters with this approach is the fact that the BAN estimator so constructed is written in terms of first and second partials of the density function, and thus involves (for the Poisson-normal convolution, for example) a sum of several rational functions of terms expressible only as infinite series. Thus, while a formal representation of a BAN estimator may be given, the consequences of making necessary approximations to obtain a usable estimate are largely unknown. Of course, certain approximations are also required to produce a pseudo MLE. The very interesting question of the comparative performance of these two approximations and other competing estimators in small or moderate size samples will be the subject of a future investigation.

An iterative procedure based on the method of pseudo maximum likelihood estimation may provide more efficient estimates than the single iteration we have discussed. In a two parameter model $F_{\theta,p}$, for example, one might estimate p by the method of moments, and obtain alternately the pseudo MLE of θ , then of p , and so forth. The convolution $P(\theta)*B(N,p)$ lends itself to application of such an algorithm since both one-parameter problems are easily solved by maximum likelihood, while the two-parameter problem is not. One feature of the algorithm suggested above is that the likelihood is guaranteed to increase with each iteration - a characteristic which some numerical procedures for finding the MLE do not possess. Under regularity conditions, the algorithm should converge to the MLE, but the speed of this convergence may preclude its use.

REFERENCES

- Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimates. Journal of the Royal Statistical Society B 32, 283-301.
- Barrett, V. D. (1966). Evaluation of the maximum-likelihood estimator where the likelihood equation has multiple roots. Biometrika 53, 151-165.
- Basu, D. (1975). Statistical information and likelihood. Sankhyā A 37, 1-71.
- Basu, D. (1977). On the elimination of nuisance parameters. Journal of the American Statistical Association, 72, 355-66.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society B 26, 211-252.
- Cushing, D. H. (1973). The Detection of Fish. Oxford: Pergamon Press.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39, 1-38.
- Ferguson, T. S. (1978). Maximum likelihood estimates of the parameters of the Cauchy distribution for samples of size 3 and 4. Journal of the American Statistical Association 73, 211-13.
- Gaffey, W. R. (1959). A consistent estimator of a component of a convolution. Annals of Mathematical Statistics 30, 198-205.
- Godambe, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations. Biometrika 63, 277-84.
- Godambe, V. P. (1977). Estimating equations. Invited paper, 40th annual meeting of the Institute of Mathematical Statistics, Seattle, Washington.
- Godambe, V. P. and Thompson, M. E. (1974). Estimating equations in the presence of a nuisance parameter. Annals of Statistics 2, 568-71.
- Johnson, N. L. and Kotz, S. (1970). Distributions in Statistics; Continuous Univariate Distributions-1. New York: Wiley and Sons.
- Kalbfleisch, J. D. and Sprott, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters. Journal of the Royal Statistical Society B 31, 175-208.
- Katz, B. (1966). Nerve, Muscle and Synapse. New York: McGraw-Hill.

- Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Annals of Mathematical Statistics 27, 887-906.
- Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. Econometrika 16, 1-32.
- Samaniego, F. J. (1976). A characterization of convoluted Poisson distributions with applications to estimation. Journal of the American Statistical Association 71, 475-79.
- Samaniego, F. J. (1977). Maximum likelihood estimation for binomially distributed signals in discrete noise. Journal of the American Statistical Association, to appear.
- Samaniego, F. J. and Hannon, G. G. (1978). Two characterizations for Pascal signals in additive noise. Technical Report, Department of Mathematics, University of California, Davis.
- Sclove, S. L. and Van Ryzin, J. (1969). Estimating the parameters of a convolution. Journal of the Royal Statistical Society B 31, 181-91.
- Shonick, W. (1970). A stochastic model for occupancy related random variables in general acute hospitals. Journal of the American Statistical Association 65, 1474-1500.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 78 - 1416	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PSEUDO MAXIMUM LIKELIHOOD ESTIMATION: THEORY AND APPLICATIONS Report # 7		5. TYPE OF REPORT & PERIOD COVERED Interim
7. AUTHOR(s) Gail G. Hannon and Francisco J. Samaniego		6. PERFORMING ORG REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of California , Davis Department of Mathematics Davis, California		8. CONTRACT OR GRANT NUMBER(s) AFOSR 77-3180
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		10. PROGRAM ELEMENT, PROJECT TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE August 1978
		13. NUMBER OF PAGES 50
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Maximum likelihood estimation, convolution, asymptotics		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Let X_1, \dots, X_n be i.i.d. random variables with probability distribution $F_{\theta, p}$ indexed by two real parameters. Let $\hat{p} = \hat{p}(X_1, \dots, X_n)$ be an estimate of p other than the maximum likelihood estimate, and let $\hat{\theta}$ be the solution of the likelihood equation $\partial/\partial\theta \ln L(x, \theta, \hat{p}) = 0$ which maximizes the likelihood. We call $\hat{\theta}$ a pseudo maximum likelihood estimate of θ .		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. Abstract continued.

and give conditions under which $\hat{\theta}$ is consistent and asymptotically normal. Pseudo maximum likelihood estimation easily extends to k parameter models, and is of interest in problems in which the likelihood surface is ill-behaved in higher dimensions but well-behaved in lower dimensions. ~~We~~ ^{are examined} ~~examine~~ several signal plus noise or convolution models, which exhibit such behavior and satisfy the regularity conditions of the asymptotic theory. For specific models, a numerical comparison of asymptotic variances suggests that a psuedo maximum likelihood estimate of the signal parameter is uniformly more efficient than estimators that have been advanced by previous authors. A number of other potential applications are noted.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)